520.38161CX2

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Kazuhisa ARUGA
Serial No.:	10/615,907
Filed:	July 10, 2003
For:	DISK SUBSYSTEM
Group:	2188
Examiner:	G. Portka

SUBMISSION OF SWORN ENGLISH TRANSLATION OF PRIORITY DOCUMENT

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

July 12, 2005

Sir:

Attached is a Sworn English Translation of the priority documents submitted on even date herein for the above-referenced application. The attached is being submitted in order to perfect Applicants claim of priority.

To the extent necessary, applicants petition for an extension of time under 37 C.F.R. section 1.136. Please charge any shortage in the fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account No. 50-1417 (Case No. 520.38161CX2) and please credit any excess fees to such Deposit Account.

Respectfully submitted,

Carl I. Brundidge
Registration No. 29,621
MATTINGLY, STANGER, MALUR & BRUNDIDGE, P.C.

CIB/jdc
703/684-1120

[Document Name] Application for Patent

[Reference No.] K98011501

[Application Date] February 2, 1999

[Destination] Commissioner, Patent Office

[International Patent Classification] G06F 3/06

[Title of the Invention] DISK SUBSYSTEM

[No. of Claims] 5

[Inventor]
[Address] 2880 Kouzu,
Odawara-shi, Kanagawa-ken
Hitachi, Ltd.,
Data Storage & Retrieval Systems Division
[Name] Kazuhisa ARUGA

[Applicant for Patent]
[Identification No.] 000005108
[Name] Hitachi, Ltd.

[Agent]
[Identification No.] 100068504
[Patent Attorney]
[Name] Katsuo OGAWA

[Designation of Charge]
[Ledger No. for Prepayment] 013088
[Amount of Payment] 21000

[List of Document Submitted]
[Object Name] Specification 01
[Object Name] Drawings 01
[Object Name] Abstract 01

[Proof] Required

[Name of the Document] Specification

[Title of the Invention]

DISK SUBSYSTEM

[Claims]

[Claim 1]

A disk subsystem comprising: a plurality of disk drive units for storing data; and a disk array controller to control data input/output from/to the disk drive units and a host computer, the disk array controller being connected to the disk drive units with a fiber channel,

characterized in that the disk array controller is switch-connected to the disk drive units.

[Claim 2]

A disk subsystem comprising: a plurality of disk drive units for storing data; and a disk array controller to control data input/output from/to the disk drive units and a host computer,

characterized in that a switch and a switch controller to switching-control the switch are provided between the disk drive units and the disk array controller, and a protocol controller is provided between the switch and the disk drive units and/or between the disk array controller and the switch.

[Claim 3]

The disk subsystem according to claim 2, characterized in that connection between the disk array controller and the switch as well as between the switch and the disk drive units is made by use of a fiber channel, and the switch is a fiber channel fabric switch.

[Claim 4]

A disk subsystem comprising: a host interface controller to control data input/output from/to a host computer; a cache memory for temporarily storing data received by the host interface controller; a parity data generator to add parity data to the data; a plurality of disk drive units for storing the data and the parity data; and a disk array controller having a disk drive interface to write the data into the disk drive units,

characterized in that the disk drive interface is provided with a protocol controller and a switch, and the plurality of disk drive units are switch-connected.

[Claim 5]

A disk subsystem comprising: a host interface controller to control data input/output from/to a host computer; a cache memory for temporarily storing data received by the host interface controller; a parity data generator to add parity data to the data; a plurality of disk drive units for storing the data and the parity data; and a disk array controller having a disk drive interface to write the data into the disk drive units,

characterized in that the disk array controller is connected to the disk drive units by use of a fiber channel, and a fabric switch having: a first protocol controller, connected to the disk drive interface, to detect ID numbers of disk drive units as subjects of access and control a fiber channel protocol; a switch controller holding the ID numbers of the respective disk drive units, to set switches in accordance with the ID numbers; and a second protocol controller, connected to the disk drive units 4, to allocate the ID numbers to the disk drive units, is provided between the disk array controller and the disk drive units.

[DETAILED DESCRIPTION OF THE INVENTION]

[0001]

[Technical field of the Invention]

The present invention relates to an electronic device including a computer system incorporating a disk subsystem, a disk array, or a disk drive. The present invention also relates to a technology, which allows high-speed transfer by means of arrayed disks connected by a fabric switch.

[0002]

[Prior Art]

In general, the connection between a disc controller device and a plurality of disk drives in a disk array may be achieved, as disclosed in the Japanese Published Unexamined Patent Application No. Hei 10-171746, by an SCSI interface or by a fiber channel arbitrated loop topology.

[0003]

The SCSI interface, which uses a time-divided data transfer method on one transfer line, negotiates with its initiator one to one for one moment on one transfer line for an access.

[0004]

The fiber channel arbitrated loop topology, on the other hand, may connect, with the SCSI interface, the initiator and disk drives in a loop by means of a serial interface, to enable time-division transfer of the data divided into frames to allow a number of communications with a plurality of devices at the same time and to allow up to 126 disk drive devices to be connected.

[0005]

[Problems to be solved by the invention]

Disk drives will become more and more compact and higher density implementation thus will ultimately realize the use of more disk drive devices.

[0006]

An SCSI interface in the Prior Art adopts a one-to-one data transfer scheme for one moment in one transfer line, which may be a drawback if one wishes to implement simultaneous communications between an initiator and a plurality of disk drives. The number of connectable disk drive units in one bus is also limited to 7 or 15. When one connects a number of drive units for one-to-one negotiation on the SCSI interface, a plurality of interfaces are required, causing difficulty in mounting. Because the number of the connectable disk drive units in one controller is so limited, one may encounter the necessity to add some further controller for connecting a large number of units to a system.

[0007]

On the other hand, in the case of fiber channel, as the disk drive has a different protocol from that of the controller, switch connection cannot be made but loop connection is made using a fiber channel arbitrated loop where many disk drives share one loop. Accordingly, if the number of disk drives connected to the same loop is increased, the data transfer speed of the disk drives is higher than the loop maximum data transfer speed. As a result, transfer cannot be performed with efficiency equal to or higher than the loop maximum data transfer speed, but can be performed at a data transfer speed approximately equivalent to that of the SCSI interface.

[0008]

[Means to solve the problems]

To solve the above problem, the present invention provides a protocol controller between a fiber channel fabric switch and disk drives for switch connection between the disk drives and a controller.

[0009]

[DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS]

A detailed description of a preferred embodiment of an external storage device (a disk subsystem) embodying the present invention will now be given referring to the accompanying drawings. Fig. 1 shows an overview of the device.

[0010]

In the external storage device shown in the figure, N disk array controllers (controller section) (1-1) to (1-N) (controllers in middle such as 1-2 are not shown, this applies to hereinbelow) are connected to a host computer (not shown) in an upper side, and provide M disk drive interface (disk drive I/F) controllers (2-1) to (2-M) in a bottom side. The hardware configuration of the disk array controller will be described below in greater details. Each of M controllers of fiber channel fabric switch (3-1) to (3-M) is respectively connected to the disk drive interface (I/F) controllers (2-1) to (2-M) for controlling disk drive units through their fiber channel interface 5. L disk drive units are connected to one fiber channel fabric switch controller, a total of M by L disk drive units (4(1,1) to 4(M,L)) are connected to the fiber channel fabric switch controllers (3-1) to (3-M) through fiber channel interfaces 6.

[0011]

Each of disk drive interface controllers (2-1) to (2-M) and disk drive units 4(1,1) to 4(M,L) has its unique identifier

(ID number) for storing data. The fiber channel fabric switch controllers (3-1) to (3-M) receive the ID numbers of the disk drive units to be connected from the disk drive interface controllers (2-1) to (2-M), to establish one-to-one connection between the corresponding disk drive interface controllers (2-1) to (2-M) and the disk drive units 4(1,1) to 4(M,L).

[0012]

Fig. 2 shows a hardware configuration of the disk array controllers (1-1) to (1-N). Data transferred thereto from the host computer (not shown) is temporarily stored in a cache memory 8 controlled by a host interface controller 7 so as to be added with parity data by a parity data generator 9, then to be split into (a total of M segments of) data blocks and parity data block(s). These data blocks and parity block(s) will be stored to a respective disk drive group (not shown) by the disk drive interface controllers (2-1) to (2-M), which are corresponding interfaces.

[0013]

To transfer data to the host computer, if there is data to be transferred thereto in the cache memory 8, then the data in the cache will be transferred to the host computer by the host interface controller 7. If the data to be transferred to the host computer is not in the cache memory 8, then the disk drive interface controllers (2-1) to (2-M) will read split data segments out of the disk drive group, concatenate split data segment blocks in the parity data generator 9, and store the complete data temporarily in the cache memory 8, and the host interface controller 7 will transfer the data to the host computer.

[0014]

The foregoing embodiment depicts a data storage method in a case of a RAID system. However, data may also be stored without the RAID system. Without the RAID system, parity data generator 9 does not exist. The data transferred from the host computer (not shown) is temporarily stored in the cache memory 8 by the host interface controller 7 and then written to any one of disk drive units in the disk drive group. When mirroring, identical data will be written into the plural disk drive units. For reading out, the data will be read out of the disk drive units, stored temporarily in the cache memory 8 and the host interface controller 7 will transfer it to the host computer.

[0015]

It should be noted that in the following description, another embodiment of disk subsystem using the RAID system will be described, however the embodiment may equivalently be made without using the RAID system.

[0016]

Fig. 3 shows a hardware configuration of the fiber channel fabric switch controllers (3-1) to (3-M). A protocol controller 16 (first protocol controller) connected to the disk driver interface controller (2-1) detects the ID numbers of the disk drive units 4(1,1) to 4(1,L) as the subjects of access and controls a fiber channel protocol. A protocol controller 16' (second protocol controller) connected to the disk drive units 4(1,1) to 4(1,L) allocates the ID numbers to the disk drive units 4(1,1) to 4(1,L), and notifies a switch controller 17 of the ID numbers of the disk drive units 4(1,1) to 4(1,L). The switch controller 17, holding the ID numbers of the respective disk drive units 4(1,1) to 4(1,L), sets switches 18 in accordance with the received ID numbers with the disk drive interface controllers (2-1) to (2-M), thus establishes one-to-one connection.

[0017]

In addition, the protocol control may be set so as to be performed in the protocol controller 16', or may be set so as to switch the protocol controller 16 with the protocol controller 16' for the data transfer from the host computer and for the data transfer to the host computer, or for the data transfer for a normal operation and for the data transfer for an operation in a disk failure.

[0018]

Another configuration may also be used in which one of the protocol controller 16 and the protocol controller 16' is used, in such a case an ID number detector means may be provided instead of the protocol controller 16, or an ID number allocating means may be provided instead of the protocol controller 16'.

[0019]

Alternatively, a protocol controller and switches may be provided within the disk drive interface controllers (2-1) to (2-M) to allow direct connection to the disk drive units 4(1,1) to 4(1,L), instead of proprietary fiber channel fabric switches provided independently in the system.

[0020]

Fig. 4 shows an operation of the fiber channel fabric switch controllers (3-1) to (3-M).

[0021]

The disk array controller (1-1) stores data split to M segments into a disk drive group (10-1). The disk drive interface controllers (2-1) to (2-M) in the disk array controller (1-1) send the ID number of disk drive units belonging to the disk drive group (10-1) to the fiber channel fabric switch controllers (3-1) to (3-M) so as to establish a switching. The protocol controller 16 in the fiber channel fabric switch controllers (3-1) to (3-M) (see Fig. 3) detects the ID number sent to request the switch controller 17 to switch the switch connection in order to achieve the protocol control pertinent to the disk drive units. The switch controller 17 (see Fig. 3) switches a switch 18 (see Fig. 3) so as to connect the disk array controller (1-1) requesting connection to the requested disk drive unit 4 belonging to the disk drive group (10-1).

[0022]

It should be recognized that since the disk array controller (1-1) is correspondingly connected to one disk drive group (10-1) through the fiber channel fabric switch controllers (3-1) to (3-M), another disk array controller (1-N) and the disk drive group (10-2) may separately perform another data transfer. Even when the disk array controller (1-N) establishes a connection to the disk drive group (10-L), the connection between the disk array controller (1-1) and the disk drive group (10-1) and the connection between the disk array controller (1-N) and the disk drive group (10-L) can operate separately from each other to perform the data transfer at a maximum data transfer rate possible between each disk array controller and respective disk drive unit.

[0023]

Although not described in this specification, the switch controller 17, when switching the connection as described above, may effectively maintain the maximum transfer window by switching the connection of the switch 18 upon reception of signals indicating that the disk drive unit connected thereto becomes ready to read/write at a time of data read or data write.

[0024]

Fig. 5 shows another extended embodiment in accordance with the present invention. In the embodiments above, the protocol controller 16 in a fiber channel fabric switch controller 3 was connected one to one to the disk drive unit 4. In the present embodiment, however, the same section is configured such that the protocol controller 16 is connected in loop to the plural disk drive units 4 through a fiber channel arbitrated loop controller 11. In this manner, an array of a plurality of inexpensive disk drive units 4 may operate at the performance level equivalent to an expensive large disk drive unit of the same capacity. In this configuration, not all disk drive units are connected in loop. Apparently the fiber channel arbitrated loop controller 11 and the plural disk drive units 4 form the single disk drive unit 4, and therefore the performance of accessing will not be degraded.

[0025]

Although not shown in the figure, if the maximum data transfer rate of the fiber channel interface is sufficiently higher with respect to the accessing speed of disk drive, the number of disk drive units 4 may be increased without aggravation of access performance, by connecting the plural disk drive units 4 to the fiber channel arbitrated loop controller 11, by connecting the plural disk drive units in the same loop, and by sharing the maximum transfer rate of the fiber channel with the plural disk drive units 4.

[0026]

Fig. 6 shows a hardware configuration of the fiber channel arbitrated loop controller 11 used for the embodiment shown in Fig. 5.

[0027]

A fiber channel arbitrated loop controller 11 comprises a loop bypass circuit 13, a plurality of disk drive unit attaching ports 12, and a fabric switch connector port 15. From disk drive units 4, a loop bypass circuit switching signal 14 is output, allowing ports to be bypassed in case of failure, then loops will keep alive, other operating disks will not be affected, and the failed disk drive unit can be detached and/or new disk drive units can be added.

[0028]

Fig. 7 shows another extended embodiment in accordance with the present invention.

[0029]

The present embodiment comprises spare disk drive unit controllers 19 each connected to respective fiber channel fabric switch controllers (3-1) to (3-M), and a plurality of spare disk drive units (4-a) and (4-b) each connected to the spare disk drive unit controllers 19. In the fiber channel fabric switch controller 3, a protocol controller 16' (see Fig. 3) connected to a disk drive group (disk drive group 4(1,2) in the figure) including a failed disk drive unit 4 is connected to a protocol controller 16' connected to the spare disk drive unit controller 19 via the switch 18. In a case where any disk drive fails, the disk array controllers (1-1) to (1-N) reconstruct data in the spare disk drive unit (4-a) or (4-b).

[0030]

In a case were an error frequently occurs in a particular disk drive 4 and there is a possibility of failure, data in the error-occurred disk drive unit 4 is transferred to the spare disk drive unit (4-a) or (4-b) and reconstructed. In a case where the disk drive unit 4 is completely broken and the data cannot be transferred, the lost data is regenerated with the cache memory 8 and the parity data generator 9 by use of data in the disk drive group including the broken disk drive unit 4, and written into the spare disk drive unit (4-a) or (4-b).

[0031]

Otherwise, the spare disk drive unit controller 19 may independently perform. For this purpose, the spare disk drive unit controller 19 includes a cache memory and a parity data regenerator. In a case were the disk drive unit 4 is completely broken, the remaining data in the disk drive group is read by the spare disk drive unit controller 19, then the lost data is regenerated and written into the spare disk drive unit (4-a) or (4-b).

[0032]

Accordingly, access from the respective disk drives holding divided data including the parity data to the spare disk drive unit controller 19 for recovery of the broken disk drive unit 4 or the failed portion, and data access from the disk drive unit 4 (the disk drive group including 4(1,1) and 4(1,L) in the figure) via the disk array controllers (1-1) to (1-N) and from the host computer, can be independently performed, thereby data reconstruction can be performed without any influence on data access from the host computer.

[0033]

In a similar manner, when a failed disk drive unit has been hot-swapped with a disk drive unit off the shelf, the recovery of the failed disk drive unit may be achieved without affecting any access from the host computer, as the spare disk drive unit controller 15 may establish one-to-one connection for the fiber channel fabric switch controllers (3-1) to (3-M), switched from the spare disk drive units (4-a) and (4-b) to a healthy disk drive unit newly hot-swapped with a failed disk drive unit, to perform data copy/recovery independently of the access from the disk array controller (1-1) to (1-N) to the disk drive group (10-1) to (10-L) (see Fig. 4).

[0034]

[Effect of the Invention]

The present invention provides the connectivity of the plural disk drive units to a disk drive unit interface without compromising the transfer performance by use of the fiber channel interface, which is a scheme of serial interface, and by applying a fiber channel fabric topology, which allows hot swapping of connectivity. The present invention further provides a solution of controlling the plural disk drive units with one or a few disk drive unit controllers, by dynamically switching the connectivity for each controller and disk drive group. In addition, the present invention provides improved reliability of the system by performing the operation of data recovery in case of disk drive unit failure, independently of the data transfer between the disk drive interface controllers and the disk drive units.

[BRIEF DESCRIPTION OF THE DRAWINGS]

[FIG. 1]

An overview of the embodiment of the present invention.

[Fig. 2]

A detailed block diagram of the disk array controller.

[Fig. 3]

A detailed block diagram of the fiber channel fabric switch controller.

[Fig. 4]

A connection block diagram of the fiber channel fabric switch.

[Fig. 5]

A connection block diagram of the fiber channel fabric switch and the arbitrated loop.

[Fig. 6]

A detailed block diagram of the fiber channel arbitrated loop controller.

[Fig. 7]

A connection block diagram of the spare disk drive unit controller.

[EXPLANATION OF NUMERALS]

1… Disk array controller

2… Disk drive interface controller

3… Fiber channel fabric switch controller

4… Disk drive unit

5… Fiber channel interface

6… Fiber channel interface

7… Host interface controller

8… Cache memory

9… Parity data generator

10… Disk drive group

11… Fiber channel arbitrated loop controller

12… Disk drive unit attaching port

13… Loop bypass circuit

14… Loop bypass circuit switching signal

15… Fabric switch connector port

16… Protocol controller

17… Switch controller

18… Switch

19… Spare disk drive unit controller.

[Name of the Document] Abstract

[Abstract]

[Problem]

In connection between a disk controller and a disk drive unit, an interface using SCSI is in a main stream, however, in a case where the number of disk drive units is increased, in order to perform one-to-one connection with one interface, many interfaces are required since switch connection cannot be performed with an existing disk drive unit using a fiber channel. This causes difficulties in mounting.

[Means of solution]

A fiber channel fabric switch controller 3 is provided between disk drive units 4 and a disk drive interface controller 2, and a protocol controller 16 is provided between a switch 18 in the fiber channel fabric switch controller 3 and the disk drive units.

[Selected Drawing] Fig. 3

FIG. 1

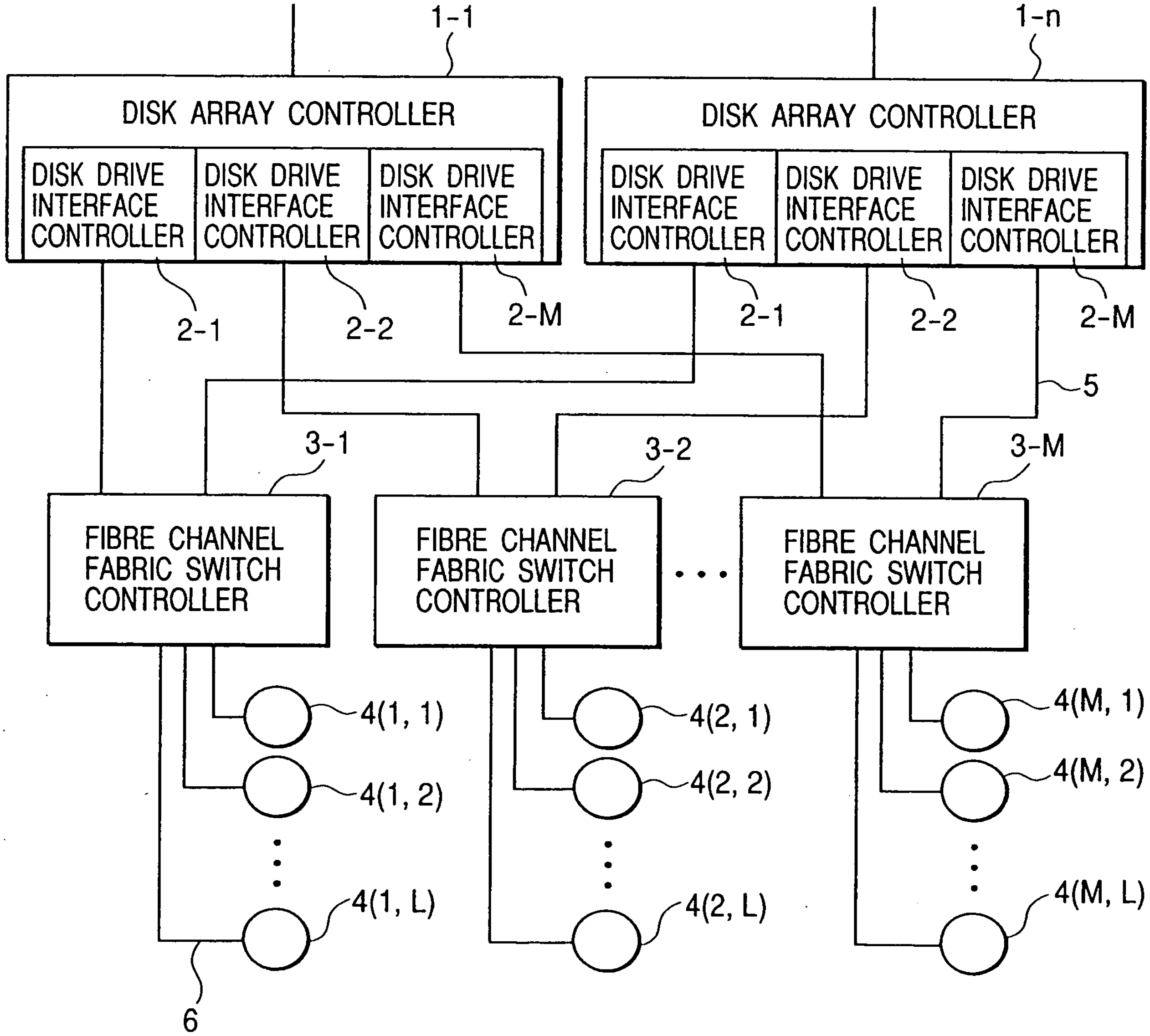

1-1
DISK ARRAY CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
1-n
DISK ARRAY CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
2-1
2-2
2-M
2-1
2-2
2-M
5
3-1
3-2
3-M
FIBRE CHANNEL FABRIC SWITCH CONTROLLER
FIBRE CHANNEL FABRIC SWITCH CONTROLLER
FIBRE CHANNEL FABRIC SWITCH CONTROLLER
4(1, 1)
4(1, 2)
4(1, L)
4(2, 1)
4(2, 2)
4(2, L)
4(M, 1)
4(M, 2)
4(M, L)
6

FIG. 2

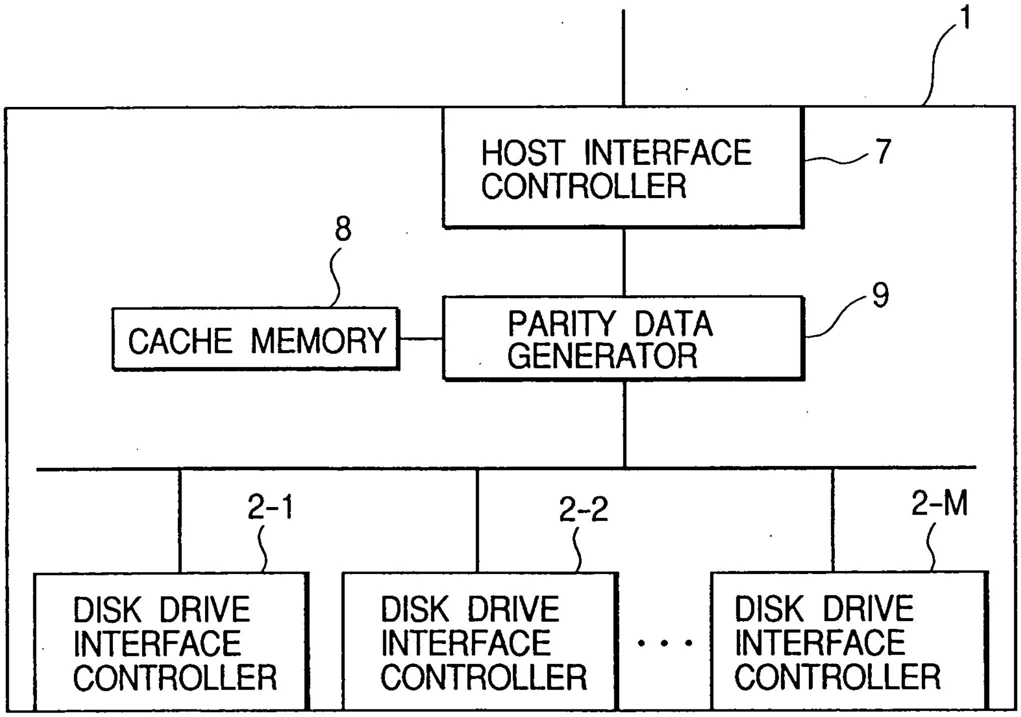

1
HOST INTERFACE CONTROLLER
7
8
CACHE MEMORY
PARITY DATA GENERATOR
9
2-1
2-2
2-M
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
. . .
DISK DRIVE INTERFACE CONTROLLER

FIG. 3

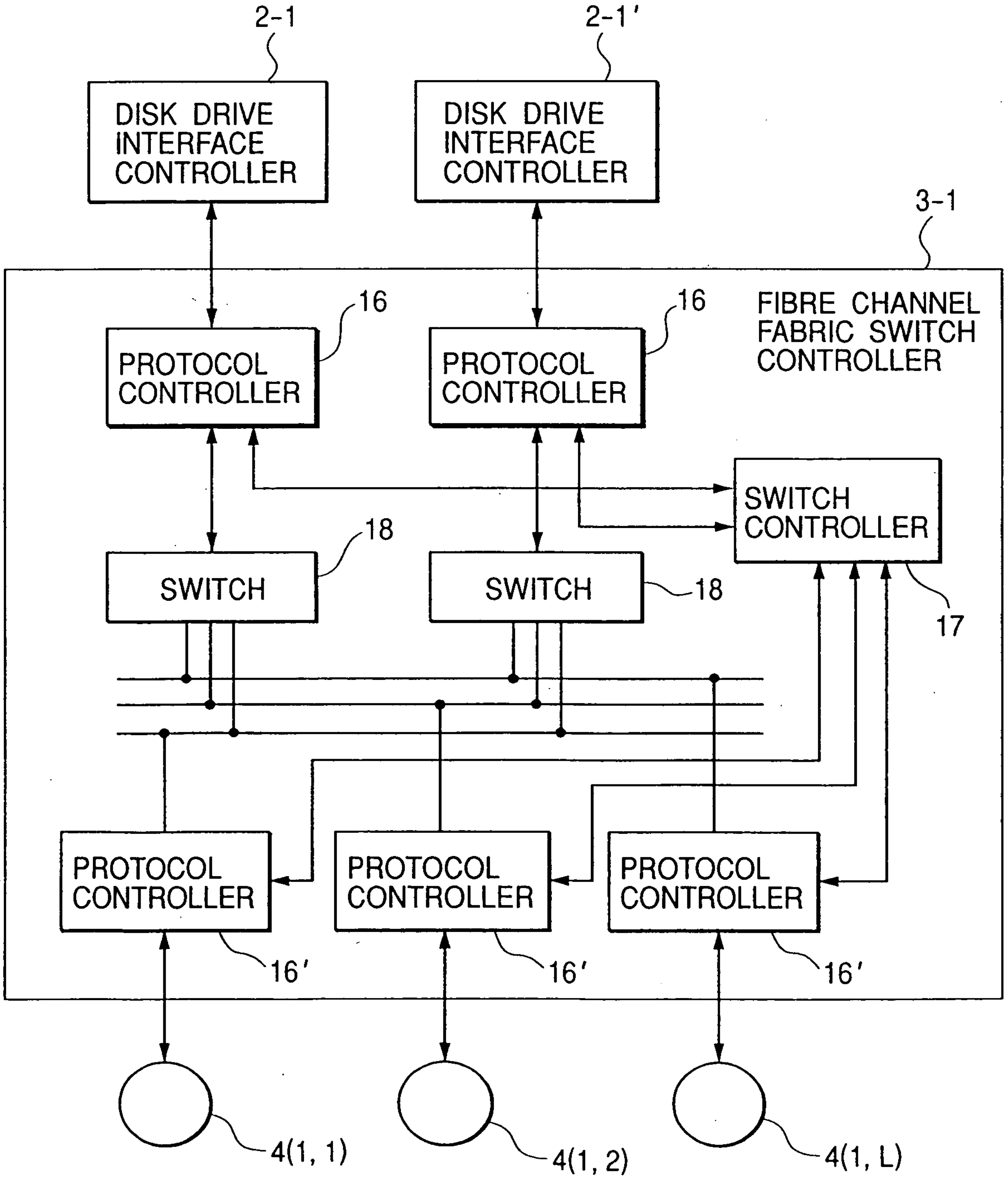

2-1
2-1'
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
3-1
FIBRE CHANNEL FABRIC SWITCH CONTROLLER
16
PROTOCOL CONTROLLER
16
PROTOCOL CONTROLLER
SWITCH CONTROLLER
17
18
SWITCH
SWITCH
18
PROTOCOL CONTROLLER
PROTOCOL CONTROLLER
PROTOCOL CONTROLLER
16'
16'
16'
4(1, 1)
4(1, 2)
4(1, L)

FIG. 4

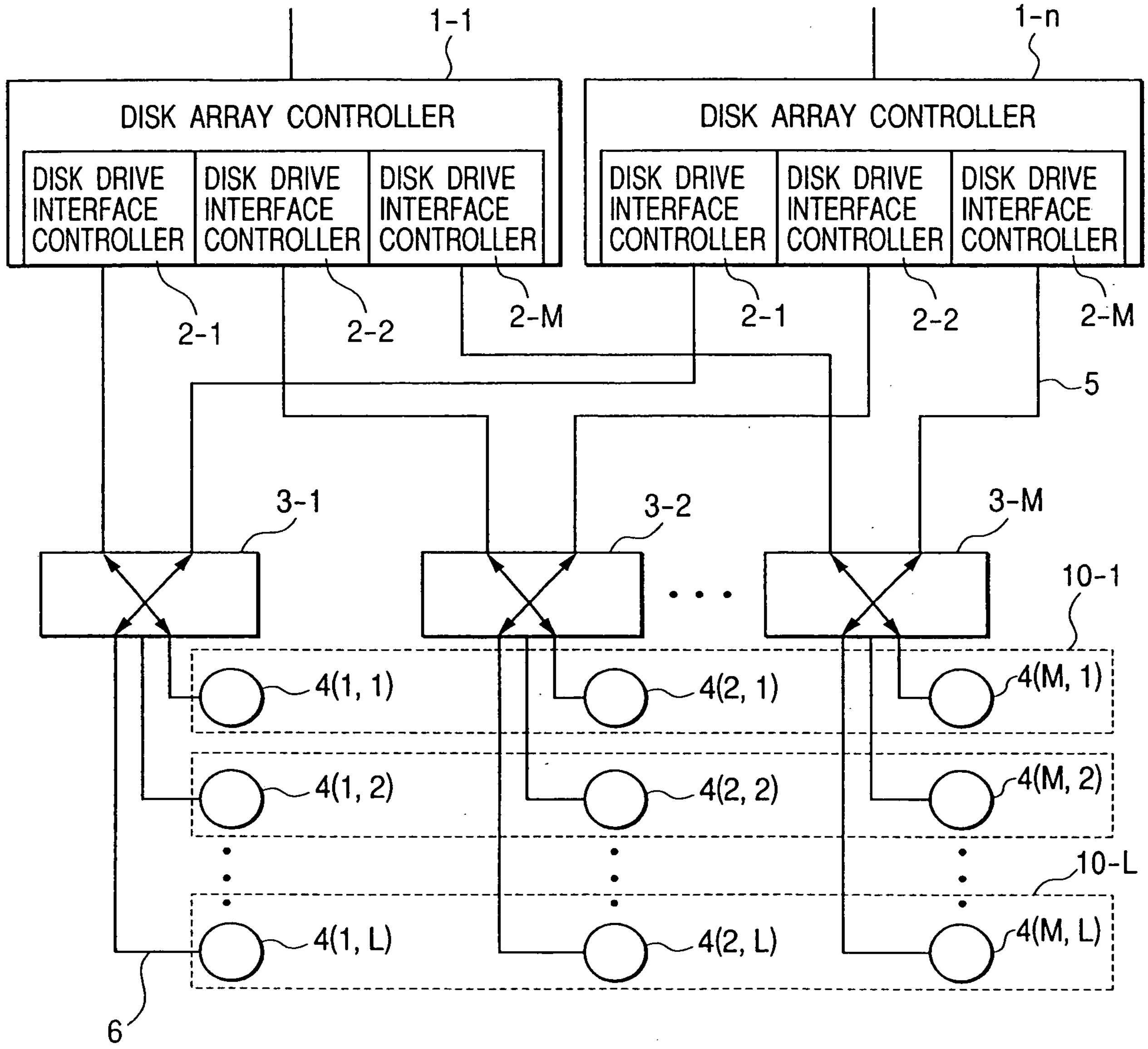

1-1
DISK ARRAY CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
1-n
DISK ARRAY CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
2-1
2-2
2-M
2-1
2-2
2-M
5
3-1
3-2
3-M
10-1
4(1, 1)
4(2, 1)
4(M, 1)
4(1, 2)
4(2, 2)
4(M, 2)
10-L
4(1, L)
4(2, L)
4(M, L)
6

FIG. 5

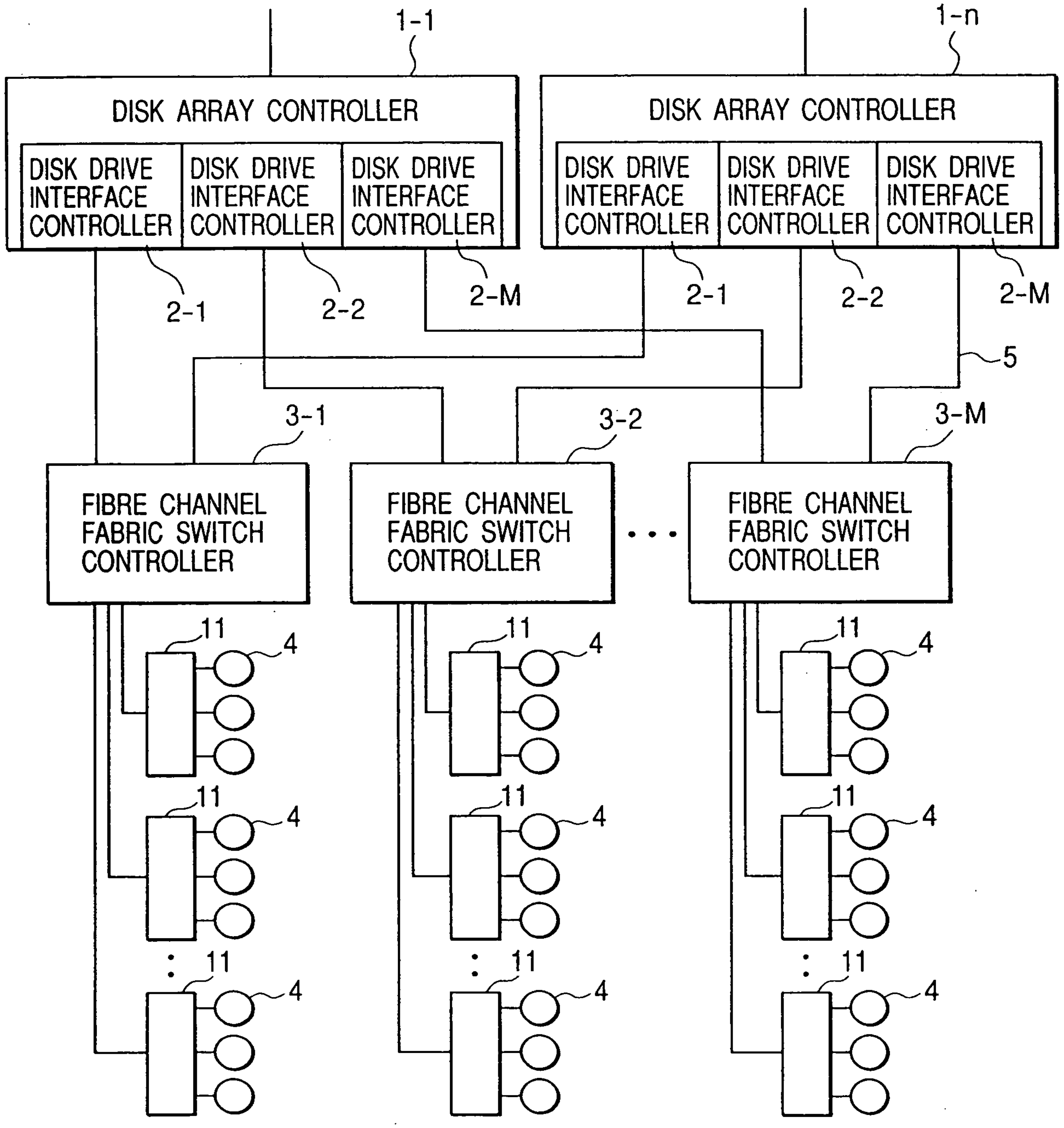

1-1
1-n
DISK ARRAY CONTROLLER
DISK ARRAY CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
2-1
2-2
2-M
2-1
2-2
2-M
5
3-1
3-2
3-M
FIBRE CHANNEL FABRIC SWITCH CONTROLLER
FIBRE CHANNEL FABRIC SWITCH CONTROLLER
FIBRE CHANNEL FABRIC SWITCH CONTROLLER
11
4
11
4
11
4

FIG. 6

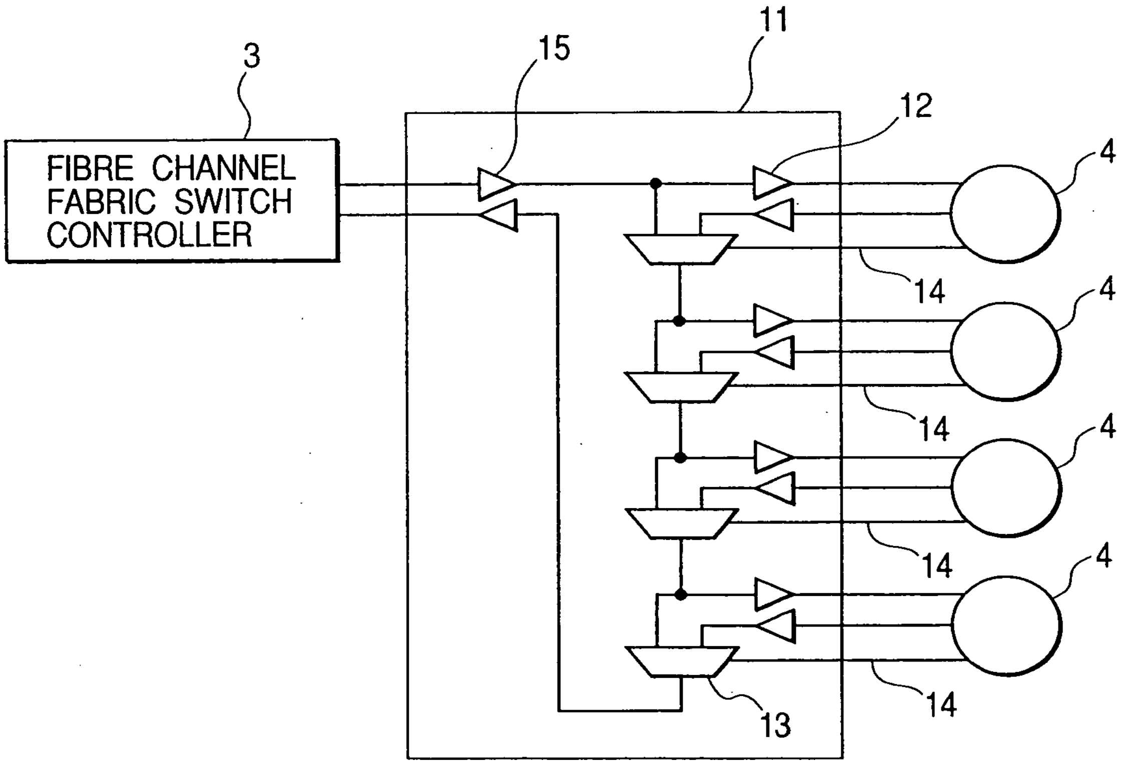
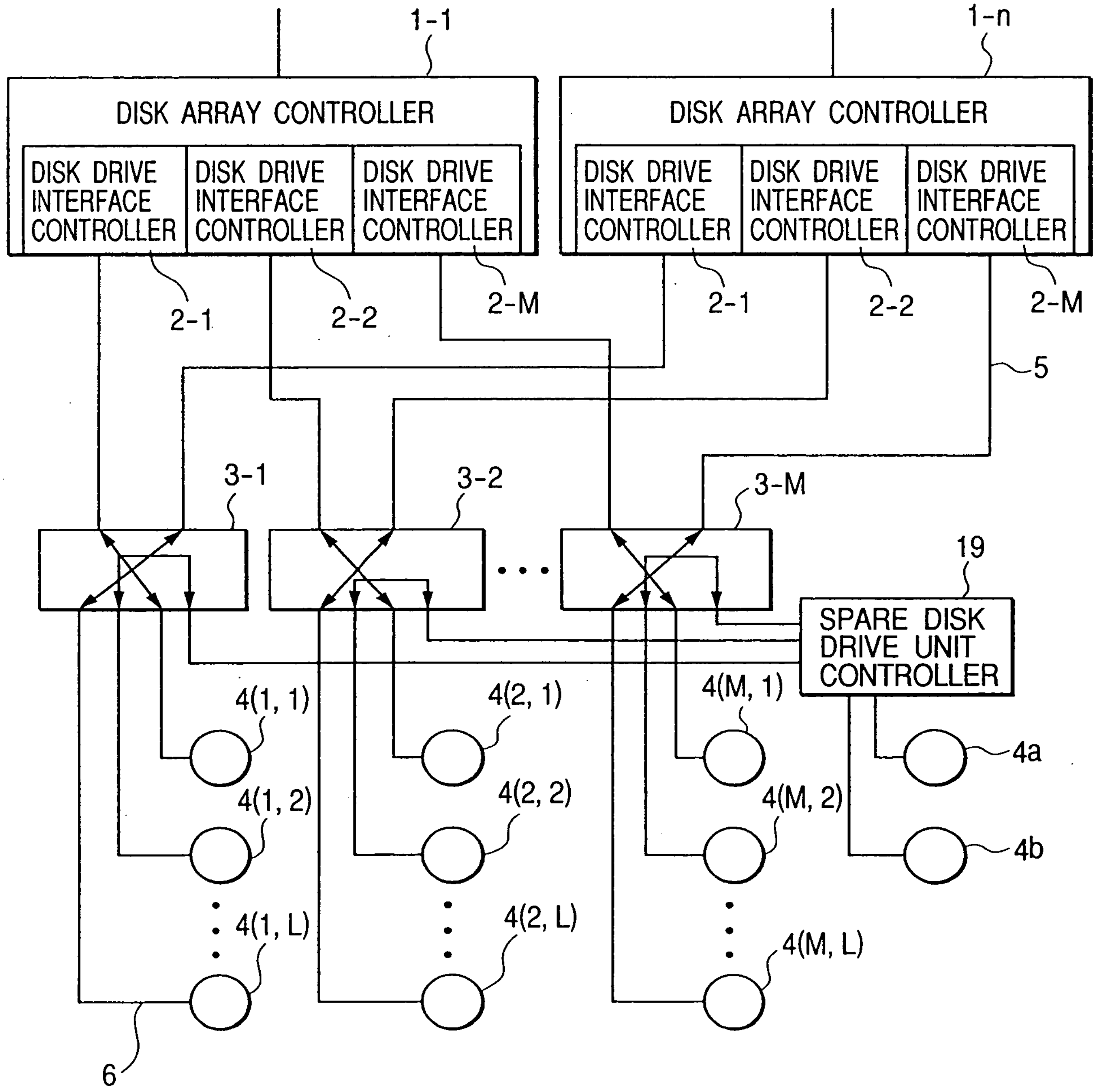

3
FIBRE CHANNEL FABRIC SWITCH CONTROLLER
15
11
12
4
14
4
14
4
14
4
13
14

FIG. 7


1-1
1-n
DISK ARRAY CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK ARRAY CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
DISK DRIVE INTERFACE CONTROLLER
2-1
2-2
2-M
2-1
2-2
2-M
5
3-1
3-2
3-M
19
SPARE DISK DRIVE UNIT CONTROLLER
4(1, 1)
4(2, 1)
4(M, 1)
4a
4(1, 2)
4(2, 2)
4(M, 2)
4b
4(1, L)
4(2, L)
4(M, L)
6